Moving Object Detection Algorithm Using Gaussian Mixture Model and SIFT Keypoint Match

Hang Dong(✉) and Xin Zhang

College of Electronics and Information Engineering,
Tongji University, Shanghai 201804, China
{dh, mic_zhangxin}@tongji.edu.cn

Abstract. In the field of image processing, Gaussian mixture model (GMM) is always used to detect and recognize moving objects. Due to the defects of GMM, there are some error detections in the final consequence. In order to eliminate the defects of GMM in moving objects detections, this paper has studied a moving object detection algorithm, combining GMM with scale-invariant feature transform (SIFT) keypoint match. First, GMM is built to obtain the distributions of background image pixels. Then, morphological processing method is applied to improve the quality of binary segmentation image and extract segmentation images of moving objects. Finally, SIFT keypoint match algorithm is used to eliminate misjudgment segmentation images by judging whether the segmentation image matches with the background template or not. Compared with original GMM, the results show that the accuracy of moving object detection has been improved.

Keywords: Moving object detection · GMM · SIFT keypoint match

1 Introduction

In the field of visual analysis, moving object detection is an important and popular research topic, which consists of classification of moving objects, tracking of moving objects and understanding of moving objects. There are some classical methods for dealing with problems of moving object detection, such as the optical flow method, the inter-frame difference method and the background subtraction method [1, 2].

GMM is one of the background subtraction method. Through training a part of the video data frames, GMM can generate a background image. With input video data, the background image can also be dynamically updated. Then, the foreground image can be separated by comparing the trained background image with each original image [3, 4]. However, with the influence of illumination variation, shaking of cameras and so on, the final moving objects are mixed with the static objects which should be classified into the background image. Thus, in the process of generating a background image using trained video data frames, the accuracy of moving object detection is low.

SIFT keypoint match algorithm can extract the feature of some key points in each image. These features which are invariant to image scale and rotation, have strong

X. Gu et al. (Eds.): MLICOM 2017, Part I, LNICST 226, pp. 22–29, 2018.
https://doi.org/10.1007/978-3-319-73564-1_3

adaptability to the change of illumination and the deformation of the image [5, 6]. Thus, these features can distinguish with each other and can be used as a basis to match two images.

This article combines GMM method with SIFT keypoint match algorithm to recognize the moving objects. First, the GMM method is used to extract moving objects in a traffic video roughly. Then, the morphological processing method which includes the opening operation and the closing operation is studied in each dynamic pixel in order to form the connected region of pixels of moving objects. Finally, the SIFT keypoint match algorithm is used to process the connected region and match them with the previous background template. The new combined method can reduce the error probability of recognizing moving objects and improve the performance of the whole system to extract moving objects in the traffic video. The improved GMM method combing with SIFT keypoint match can complete the extraction of moving vehicles in a complicated traffic video.

2 Conventional Method

In this part, the theories and functions of three conventional methods including GMM method, morphological processing method and SIFT keypoint match algorithm are briefly introduced as follows.

2.1 GMM Method

GMM is a combination of multiple Gaussian distributions [3, 4], which is used to describe the distributions of background pixel value in this paper. The sample is obtained by the Eq. (1):

$$\{X_1, X_2, \cdots, X_t\} = \{I(x_0, y_0, i); 1 \le i \le t\}. \tag{1}$$

In Eq. (1), X_t is the pixel value of the t frame, $\left(x_{0,}\ y_{0,}\right)$ is the position of pixel. X_t obeys the mixed distribution whose probability density function is shown in Eq. (2).

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \times \eta(X_i,\ \mu_{i,t},\ \Sigma_{i,t}). \tag{2}$$

In Eq. (2), K is the number of distribution in the GMM, which depends on the complexity of the background. In this paper, we consider the change of background of video frame as the result of superposition of multiple Gaussian distributions. The value of K are set as 3 in our model, $\omega_{i,t}$, $\mu_{i,t}$, $\Sigma_{i,t}$ are the weight, the mean and the covariance matrix of the i distribution in t moment respectively. $\eta(X_i,\ \mu_{i,t},\ \Sigma_{i,t})$ is a corresponding probability density function, which is shown in the Eq. (3).

$$\eta(X_i,\ \mu,\ \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-1/2(X_t-\mu)^T\Sigma^{-1}(X_t-\mu)}. \tag{3}$$

In Eq. (3), n is the dimension of X_t, and Σ is the covariance matrix of each dimension of X_t. The probability density function of the mixed distribution is shown in the Fig. 1.

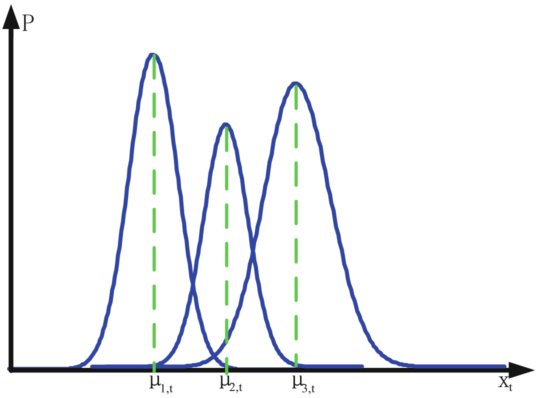

Fig. 1. The probability density function of the mixed distribution.

According to [3, 4], each distribution parameter will be updated as shown in Eq. (4) after the initialization of the GMM.

$$\begin{cases} \hat{\omega}_{i,t} = (1-\alpha)\omega_{i,t} + \alpha P(k|X_i,\ \mu_{i,t},\ \sigma_{i,t}) \\ \hat{\mu}_{i,t} = (1-\rho)\mu_{i,t} + \rho X_i \\ \hat{\sigma}^2_{i,t} = (1-\rho)\sigma^2_{i,t} + \rho(X_i-\mu_{i,t})^T(X_i-\mu_{i,t}) \end{cases} \tag{4}$$

In Eq. (4), in the t moment, $\hat{\omega}_{i,t}$ is the estimated value of $\omega_{i,t}$, $\hat{\mu}_{i,t}$ is the estimated value of $\mu_{i,t}$, and $\hat{\sigma}^2_{i,t}$ is the estimated value of $\sigma^2_{i,t}$. Those estimated values are regarded as the value of $\omega_{i,t}$, $\mu_{i,t}$, $\hat{\sigma}^2_{i,t}$ in the $t+1$ moment. α is the rate of learning, which determines the updating speed. $\rho = {}^{\alpha}\!/_{\omega_{i,t}}$ is the learning rate of parameter. The judgment of matching X_t with the k distribution is shown in Eq. (5).

$$|X_t - \mu_{k,t}| < D\delta_{k,t}\,. \tag{5}$$

If the matching can be satisfied well, $P(k|X_i,\ \mu_{i,t},\ \sigma_{i,t}) = 1$, or the value would be 0. The general value of D is 2.5. After Gaussian distribution is updated, the weight should be normalized according to $\sum_{i=1}^{K}\omega_{i,t} = 1$, and then the weight should be sorted by size of

ω/σ. Finally, first B Gaussian distributions can be selected to describe the background image in Eq. (6).

$$B = \arg\min_{\mathrm{b}} \left(\sum_{i=1}^{b} \omega_{i,t} > T \right). \tag{6}$$

2.2 Morphological Processing Method

The morphological processing method contains the opening operation and the closing operation. They both consist of corrosion and expansion. The only difference between them is the order.

Corrosion is a process of eliminating boundary points and contracting the boundary inward. It can be used to eliminate small and meaningless pixel objects. On the other hand, expansion is the process of merging all the background points. It can be used to fill holes in objects.

Opening operation is a process of expansion after corrosion. It can eliminate small objects, separate the objects in the fine points and smooth the boundary of the large object. Meanwhile, it doesn't significantly change the area. Closing operation is a process of corrosion after expansion. It can fill the body with tiny holes, connect nearby objects and smooth its boundaries. It doesn't significantly change the area either.

2.3 SIFT Keypoint Match Algorithm

The SIFT key point match algorithm was proposed by Lowe in 1999 and perfected in 2004 [5, 6]. The SIFT feature is based on the interest points of some local appearance on the object and has nothing to do with the size and rotation of the image. The SIFT feature also has a high tolerance of light, noise, and micro vision changes.

The algorithm consists of following six steps. (1) Generate Gaussian difference pyramid and construct scale space. (2) Detect the spatial extreme points. (3) Locate the key points precisely. (4) Allocate the direction information of stable key points. (5) Describe the key points. (6) Match the key points. The flowchart of SIFT keypoint match algorithm is shown in Fig. 2.

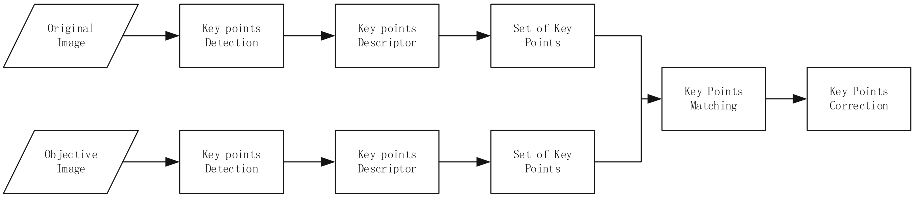

Fig. 2. The flowchart of SIFT keypoint match algorithm.

3 Method Combined GMM with SIFT Keypoint Match

The GMM is used to model the video image, which can generate the background image. The foreground image can be generated by subtracting from the original image. However, there are still misjudgments in obtaining the foreground influenced by illumination variation, partial occlusions and shaking of camera. The pixels judged to be the foreground is under the morphological processing, which means that the connected region is obtained by opening operation and closing operation. As a result, these pixels become a whole object rather than the scattered pixels. The background image of the previous GMM is regarded as a static graphic template.

Then the connected domain is surrounded by the minimum area of the rectangle, which represents the moving object. However, there are some background images in it, and the SIFT keypoint match algorithm is used to find out the right foreground by matching the background template. Through extracting rectangular image I_1, we can obtain its location, length and width. These information can be used to locate the rectangle in background template which is called I_2. The number of SIFT key points in rectangular image I_1 is N_1 and the number of SIFT key points in rectangular image I_2 is N_2. The number of matching points in rectangular image I_1 and rectangular image I_2 is M. If the relationship between N_1 and M satisfies that as shown in Eq. (7),

$$P = \frac{M}{N_1} > R,\ R = 0.7\,. \tag{7}$$

then, the rectangular image I_1 and the rectangular image I_2 will be considered as matching successfully, which means that the rectangular image I_1 is a static image rather than a moving object. Thus, the rectangular image I_1 should be removed from the foreground. The flowchart of the method is shown in Fig. 3.

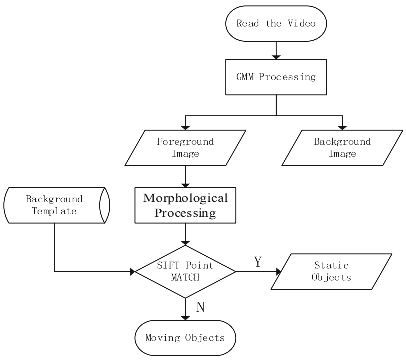

Fig. 3. The flowchart of the method.

4 Results and Analysis

The improved algorithm is compiled by using the image processing toolkit from MATLAB software. We select two videos as the testing videos [8]. Comparing with the original GMM algorithm, the experimental results show that the moving object detection accuracy rate has a great improvement when applying the improved GMM algorithm combined SIFT keypoint matching algorithm. For the video of crossroads, the compared results are as follows (Figs. 4, 5, 6 and 7).

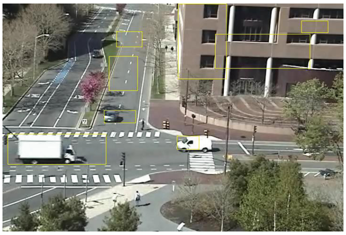

Fig. 4. The result of the detection of original GMM algorithm in video 1.

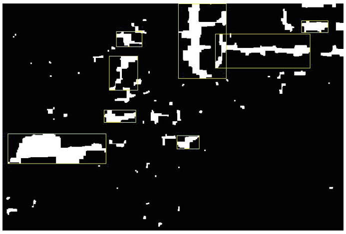

Fig. 5. The result of the binarization of foreground and background in video 1.

Fig. 6. The result of the improved algorithm in video 1.

In the case of time complexity, the proposed algorithm takes more time to deal with the video per frame than the original GMM algorithm. However, it can satisfy the need of video analysis better. Figure 8 shows the detection results of moving objects using

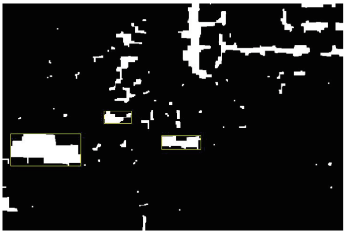

Fig. 7. The result of the binarization of foreground and background in video 1.

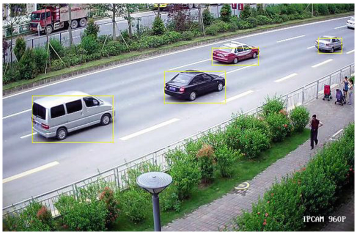

Fig. 8. The result of the improved algorithm in video 2.

the improved GMM algorithm combined with SIFT keypoint match algorithm. In the video of roads, the number of video frame is 5258 and the total number of moving objects is 85. Using the traditional algorithm, the number of wrong detection moving objects is 21 and the correct rate of detection moving objects is 72%. While using the improved algorithm, the correct number of detection moving objects is 74 and the correct detection rate of moving objects is 87%. Thus, one can find the improved algorithm has a better performance than the traditional algorithm.

5 Conclusion

In this paper, a moving object detection algorithm combined GMM algorithm with SIFT keypoint match algorithm is studied. First, the basic GMM was used to obtain the moving pixels in each video frame. The morphological processing was applied to group the rectangular image from moving pixels. Second, the SIFT keypoint match algorithm was used to obtain moving objects by distinguishing the foreground and background. Finally, the algorithm was evaluated on two different videos. It is shown that the performance of the moving objects detection algorithm is better than the traditional algorithm. Our future work will focus on how to reduce the computational complexity.

References

1. Pietikäinen, M., Hadid, A., Zhao, G., et al.: Background subtraction. Surf. Sci. **461**(1–3), 1–15 (2011)
2. Piccardi, M.: Background subtraction techniques: a review. In: IEEE International Conference on Systems, Man and Cybernetics, vol. 4, pp. 3099–3104. IEEE (2005)
3. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Detection, vol. 2, pp. 246–252. IEEE Xplore (1999)
4. KaewTraKulPong, P., Bowden, R.: An improved adaptive background mixture model for real-time tracking with shadow detection. In: Remagnino, P., Jones, G.A., Paragios, N., Regazzoni, C.S. (eds.) Video-Based Surveillance Systems, pp. 135–144. Springer, Boston (2002). https://doi.org/10.1007/978-1-4615-0913-4_11
5. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. Kluwer Academic Publishers, Dordrecht (2004)
6. Lowe, D.G.: Object detection from local scale-invariant features. In: The Proceedings of the 7th IEEE International Conference on Computer Vision. IEEE (2002)
7. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Detection, ICPR 2004, vol. 2. IEEE (2004)
8. Wang, X., Ma, X., Grimson, W.E.: Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. IEEE Trans. Pattern Anal. Mach. Intell. **31**(3), 539–555 (2009)